

阿部 賢司¹ 黒川 一滋¹ 武田 和也¹ 藤崎 博也¹ 大野 澄雄²

¹東京理科大学 ²東京工科大学

1. はじめに

音声認識において話題や話者の内部状態に適応した言語モデルを用いる場合、生起し得る単語列の候補数が減少するため、認識率は高くなることが予想される。本稿では、このような観点から、一般的なコーパスから作成した平均的な言語モデルを用いた場合を基準とし、筆者らが先に提案した知的情報検索に用いるための対話管理手法 [1,2] に基づいて構築した、話題を“学術情報検索”に限定した場合のユーザの発話の言語モデル、および、ユーザの内部状態に適応した言語モデルの採用による認識率の改善の効果を実験により検証した結果について述べる。

2. ユーザの内部状態に着目した音声認識

音声言語の生成過程を音声認識の立場からモデル化する試みは 1970 年代から行われており、F. Jelinek [3] は音響パラメータベクトルの時系列 A に対する単語列 W の事後確率 $P(W|A)$ を最大化する W を選択する問題として音声認識を定式化した。これは、現在の確率的手法による音声認識の基礎となっている。一方、B.-H. Juang [4] は言語モデルの前段階としてメッセージ源 (概念モデル) を設け、音響パラメータベクトルの時系列 A に対するメッセージ M の事後確率を最大化する問題として定式化した。この場合、言語モデルはメッセージ M に適応して変化するため、認識率の向上が期待されるが、その程度は僅少であったと報告されている [5]。

これに対して筆者らは、特定の発話の生起確率を左右する最大の要因が話者の心の状態であると考え、特に、学術情報検索におけるユーザの内部状態に依存する言語モデルを対話音声認識に適用する手法について検討した。

3. 学術情報検索を目的とした対話の管理手法

筆者らが提案した手法では、ユーザおよびシステムの内部状態を別々の確率的有限状態オートマトンとしてモデル化し、ユーザの内部状態や発話内容を推定・予測しながら対話を効率的に管理する。この手法を具体化するため、まず、学術情報検索を目的とした模擬対話 (100 対話 : 3417 発話) を収集・分析し、各発話の意図のタイプ (以下、“発話タイプ”と呼ぶ) と、それらの発話が生起する際のユーザおよびシステムの内部状態を調査した。次に、その結果に基づいて、各発話に発話タイプとユーザおよびシステムの内部状態に関する情報を付加したタグつき対話コーパスを作成し、そこから求められる状態遷移規則にしたがって、ユーザモデルおよびシステムモデルを構築した。

構築したユーザモデルを図 1 に示す (システムモデルも同様に記述することができる)。この図において、Su1 ~ Su8 はユーザの内部状態を、また u1 ~ u12 は各状態からの出力 (発話タイプ) 表す。この図からも明らかのように、各状態からの出力 (発話タイプは)

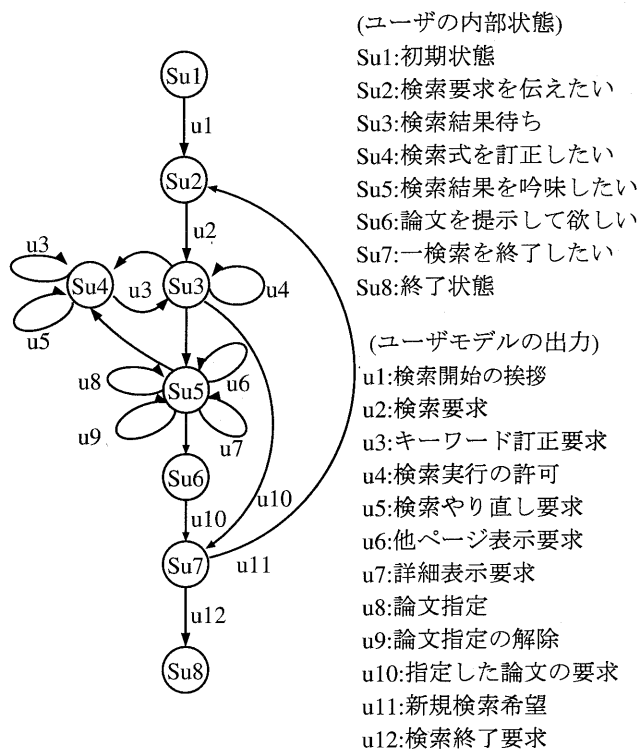


図 1. ユーザモデル

A Method for Spoken Dialogue Recognition Based on a User Model and its Evaluation

Kenji Abe¹, Kazushige Kurokawa¹, Kazunari Taketa¹, Hiroya Fujisaki¹ and Sumio Ohno²

¹Science University of Tokyo, 2641 Yamazaki, Noda, 278-8510

²Tokyo University of Technology, 1404-1 Katakura, Hachioji, 192-0982

限定されているため、これらの状態に適応した言語モデルを対話音声認識に適用することにより、生起し得る単語列の候補数が減少し、認識率は高くなることが予想される。

ここで、言語モデルとして単語バイグラムを用い、話題を“学術情報検索”に限定した場合のユーザの発話のモデルと、状態に適応したモデルとを perplexity に着目して比較した結果を表 1 に示す。表中の重み付き平均とは、各状態 Su1 ~ Su7 の出現確率に基づいて重み付けした場合の perplexity の平均値である。

いずれの場合も、状態に適応したバイグラムの方が話題を限定しただけのバイグラムよりも perplexity が低い。

表 1 単語バイグラムの perplexity

バイグラムの種類	perplexity
1) 話題に適応した単語バイグラム	4.393
2) 状態に適応した単語バイグラム:	
Su1 における単語バイグラム	1.601
Su2 における単語バイグラム	2.646
Su3 における単語バイグラム	1.810
Su4 における単語バイグラム	3.267
Su5 における単語バイグラム	3.606
Su6 における単語バイグラム	1.685
Su7 における単語バイグラム	1.665
Su1 ~ Su7 の重み付き平均	3.051

4. 対話音声認識実験

一般的なコーパスから構築した平均的な言語モデルとして、文献 [6] の基本ソフトウェアに収録されている 20000 語彙からなる単語バイグラム (毎日新聞 CD-ROM 1991 年版から 1994 年版から構築) を、また、話題に適応した言語モデルとして、先に述べた対話コーパス (収録語彙数: 307) から求めた単語バイグラムを、さらに、状態に適応した言語モデルとして、状態毎 (Su1 ~ Su7) の各発話から求めた単語バイグラムを用意し、それらの言語モデルを対話音声認識に適用したときの効果を実験的に検証した。実験では、音声認識器として Julius [6] を使用し、また、入力データとして、情報検索を目的としたユーザ発話 (60 発話) を用いた。

実験結果を表 2、3 に示す。表 2 は単語認識率に着目して比較した結果を、表 3 は対話音声理解率に着目して比較した結果を示している。ここで、対話音声理解率に関しては、入力データの意味的内容が正しく

再現されている場合を正解とした。

単語認識率、対話音声理解率のいずれにおいても、話題を限定し、さらに状態に適応した言語モデルを用いた場合の結果が最も高い値を示している。

表 2 各言語モデルを用いたときの単語認識率

言語モデル	認識率 [%]
平均的なモデル	71.8
話題に適応したモデル	73.1
状態に適応したモデル	77.1

表 3 各言語モデルを用いたときの対話音声理解率

言語モデル	理解率 [%]
平均的なモデル	61.9
話題に適応したモデル	70.0
状態に適応したモデル	75.0

5. おわりに

本稿では、対話音声認識に、話題およびユーザの内部状態に適応した言語モデルを適用する方法について検討した結果を述べ、この方法が、一般的なコーパスから構築した平均的な言語モデルを用いる従来の方法よりも有効であることを示した。なお、本稿では省略したが、ユーザモデルの自動獲得方法および内部状態の推定方法についても検討している [7]。

参考文献

- [1] H. Fujisaki, H. Kameda, S. Ohno, T. Ito, K. Tajima and K. Abe: "An intelligent system for information retrieval over the Internet through spoken dialogue," *Proceedings of Eurospeech'97*, vol.3, pp.1675-1678 (1997).
- [2] K. Abe, K. Kurokawa, K. Taketa, S. Ohno and H. Fujisaki: "A new method for dialogue management in an intelligent system for information retrieval," *Proceedings of ICSLP2000*, vol. 2, pp. 118-121 (2000).
- [3] F. Jelinek: "Continuous speech recognition by statistical methods," *Proceedings of IEEE*, vol. 64, no. 4, pp. 532-556 (1976).
- [4] B.-H. Juang: "Automatic speech recognition: Problems, progress & prospects," *IEEE Workshop on Neural Networks for Signal Processing* (1996).
- [5] 大附克年, 古井貞熙, 桜井直之, 岩崎淳, 張志鵬: "ニュース音声認識のための言語モデルと音響モデルの検討," *信学技報*, vol. 98, no. 463, pp. 1-7 (1998).
- [6] 河原達也, 李晃伸, 小林哲則, 武田一哉, 峯松信明, 嵯峨山茂樹, 伊藤克互, 伊藤彰則, 山本幹雄, 山田篤, 宇津呂武仁, 鹿野清宏: *日本ディクテーション基本ソフトウェア - 1999 年度版 -* (2000).
- [7] 阿部賢司: "An intelligent system for academic information retrieval based on human-machine spoken dialogue," *東京理科大学大学院基礎工学研究科平成 12 年度博士論文* (2001).